**Practical No 1**

**Aim: Implementing k-means classification Technique.**

**Description :-**

The algorithm will categorize the items into k groups of similarity. To calculate that similarity, we will use the euclidean distance as measurement.

The algorithm works as follows:

First, we initialize k points, called means, randomly. We categorize each item to its closest mean and we update the mean’s coordinates, which are the averages of the items categorized in that mean so far. We repeat the process for a given number of iterations and at the end, we have our clusters.

**Methods :-**

1. numpy.random.randint(low, high=None, size=None) :-

Return random integers from low (inclusive) to high (exclusive).

2. matplotlib.pyplot.figure(figsize=(x,y)) :-

Create a new figure, or activate an existing figure.

3. matplotlib.pyplot.scatter(x, y, color =’k’) :-

With Pyplot, you can use the scatter() function to draw a scatter plot.

4. matplotlib.pyplot.xlim(\*args, \*\*kwargs) :-

The xlim() function in pyplot module of matplotlib library is used to get or set the x-limits of the current axes.

5. matplotlib.pyplot.ylim(\*args, \*\*kwargs) :-

The ylim() function in pyplot module of matplotlib library is used to get or set the y-limits of the current axes.

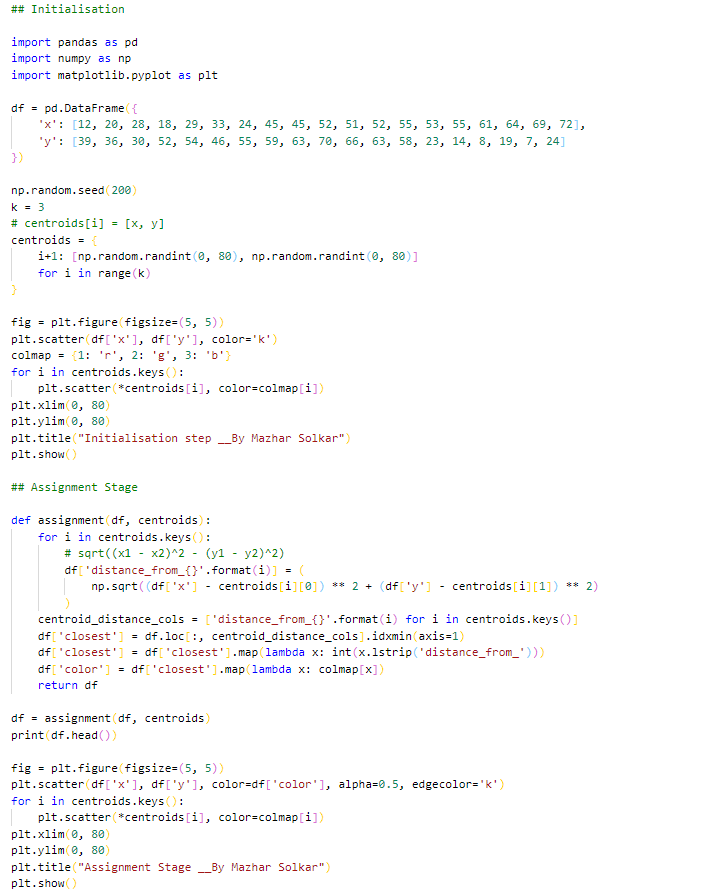
6. matplotlib.pyplot.show() :-

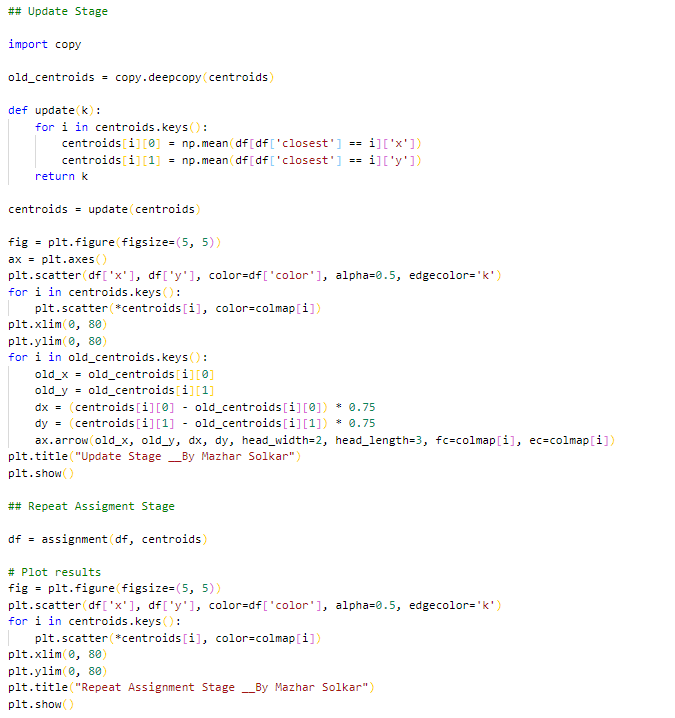
This method is used to display the graph.

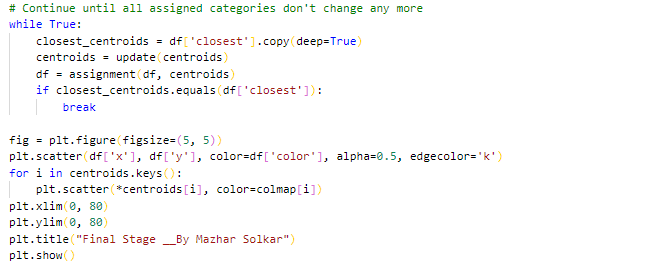
7. df.head() :-

This method is used to obtain size of the dataset.

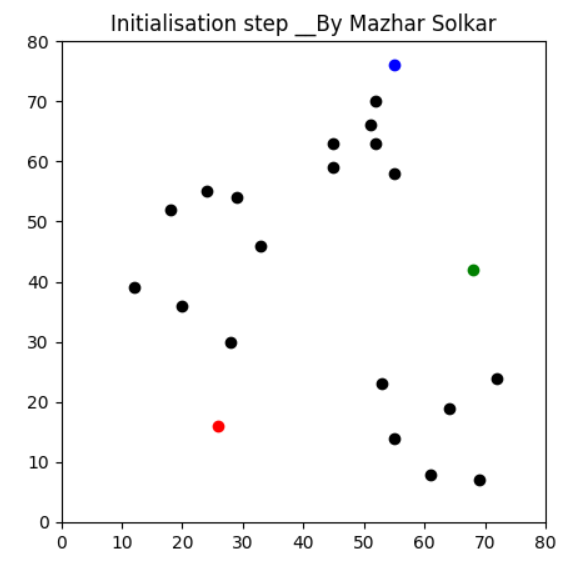
**Program :-**

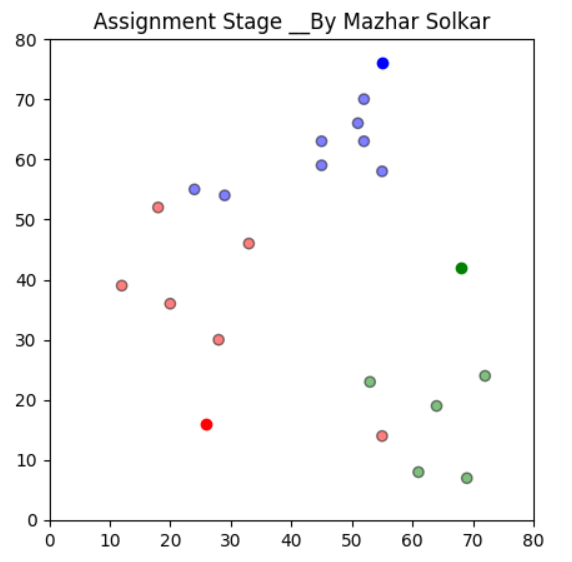


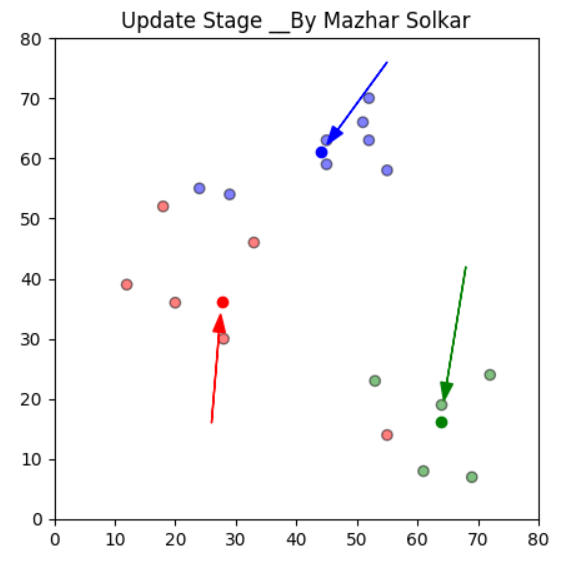


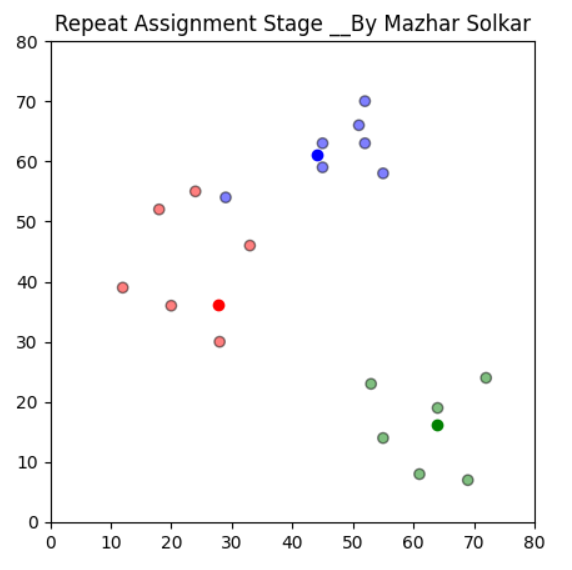


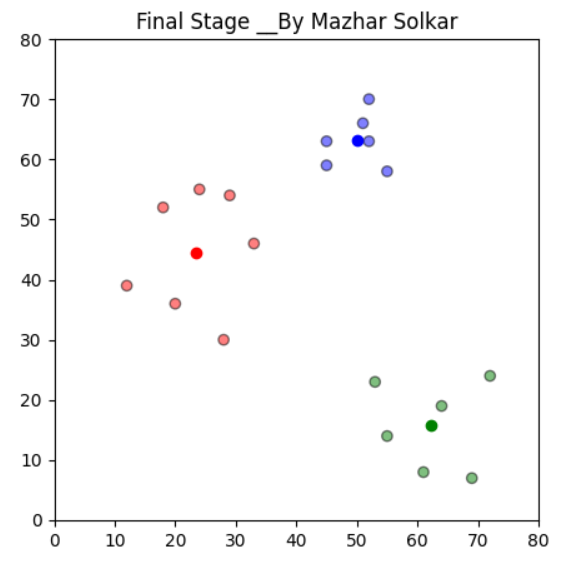
**Output :-**

****

****

****

****

****

**Practical No 2**

**Aim: Implementing Linear Regression**

**Description: -**

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

**Methods :-**

1. pd.read\_csv(inputfilename) :-

This method is used to read the csv files.

2. dataframe.iloc[:,[colno\_1,colon\_3]] :-

This method is used to fetch specific row of specific columns.

3. train\_test\_split(x,y,test\_size=0.25,random\_state=0) :-

This method is used to split dataframe into training and testing dataset.

4. StandardScaler() :-

This method is used for feature scaling.

5. SVC(kernel='linear', random\_state=0) :-

This method is used for linear support vector classifier.

6. metrics.accuracy\_score(y\_test,y\_pred) :-

This method is used to check the accuracy score.

7. model.coef\_ :-

model.coef\_ is used to obtain coefficient value and

8. model.intercept\_ :-

model.intercept\_ is used to obtain intercept value.

9. model.score(waist,weight) :-

This method is used to check the accuracy of the model.

10. model.predict(Waist\_new) :-

This method is used to predict the value based on trained dataset.

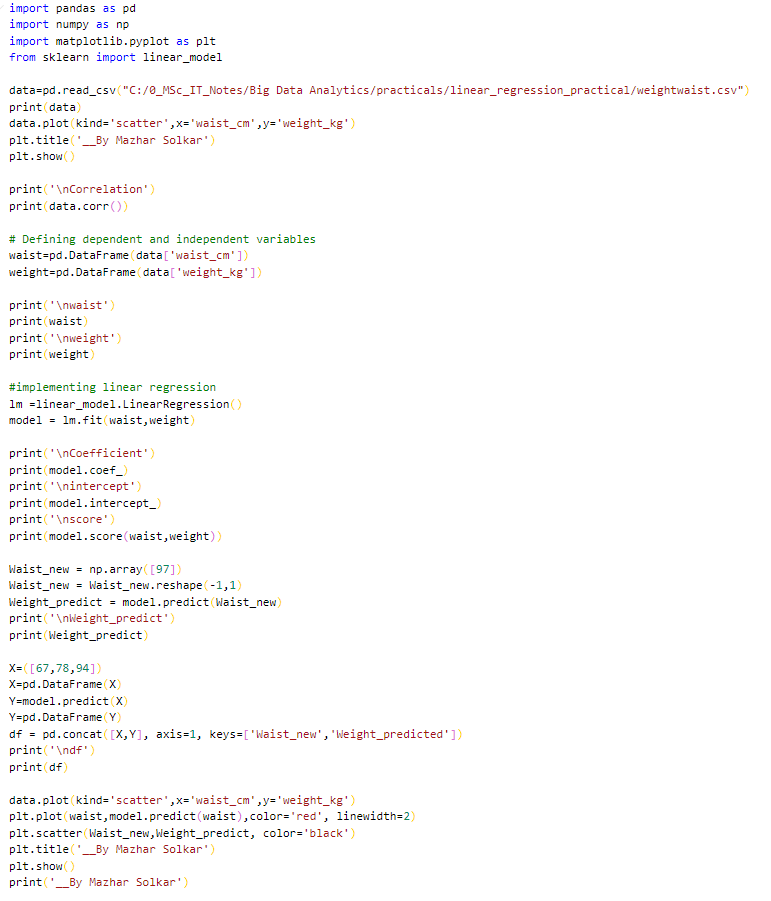
11. data.corr() :-

This method is used to obtain correlation.

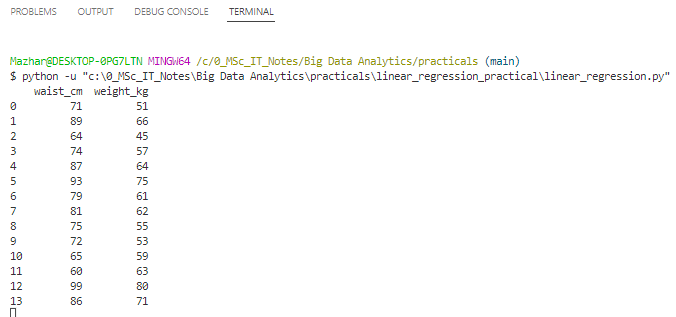
12. lm.fit(waist, weight) :-

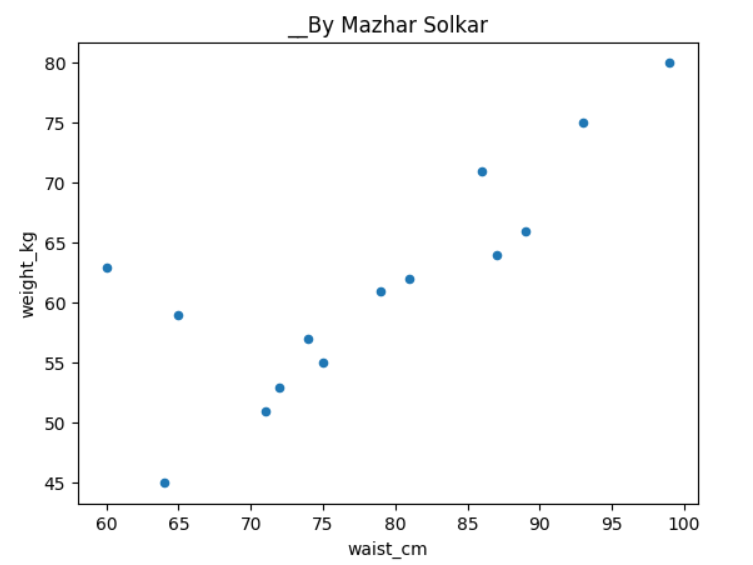
fit() is used to train model.

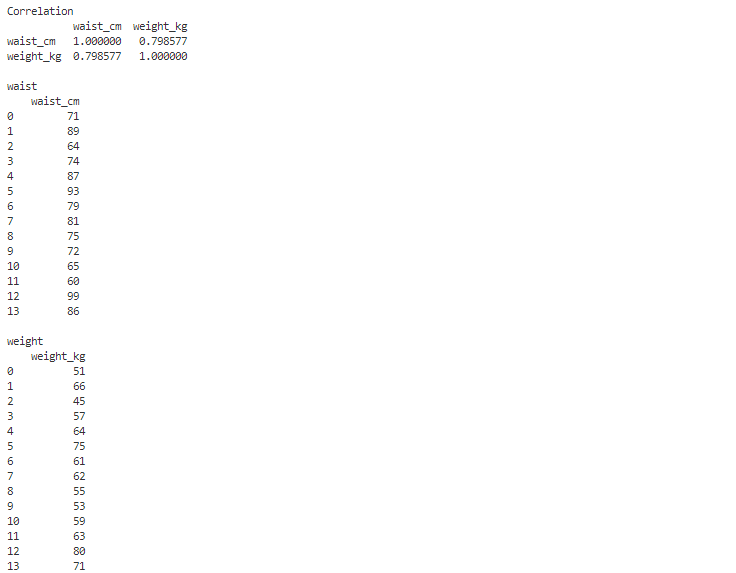
**Program :**-



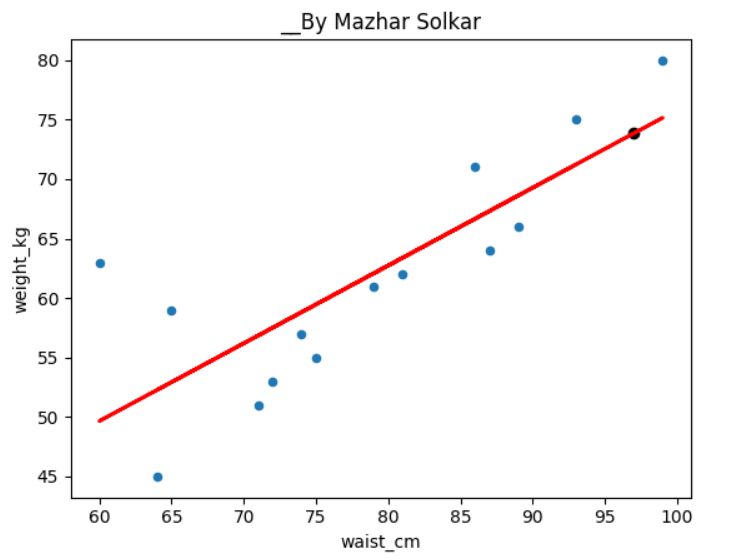
**Output :-**











**Practical No 3**

**Aim : Implementing Logistic Regression.**

**Description :-**

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

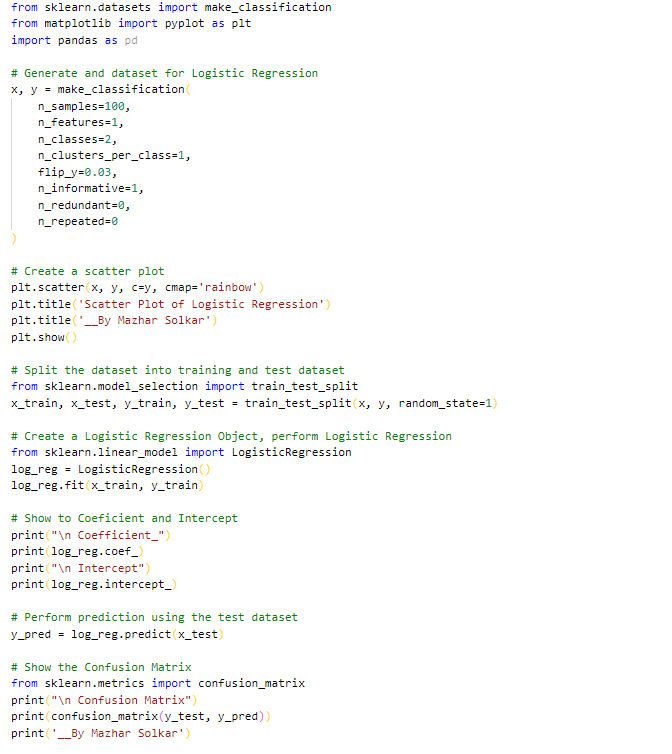
In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc. Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

**Methods :-**

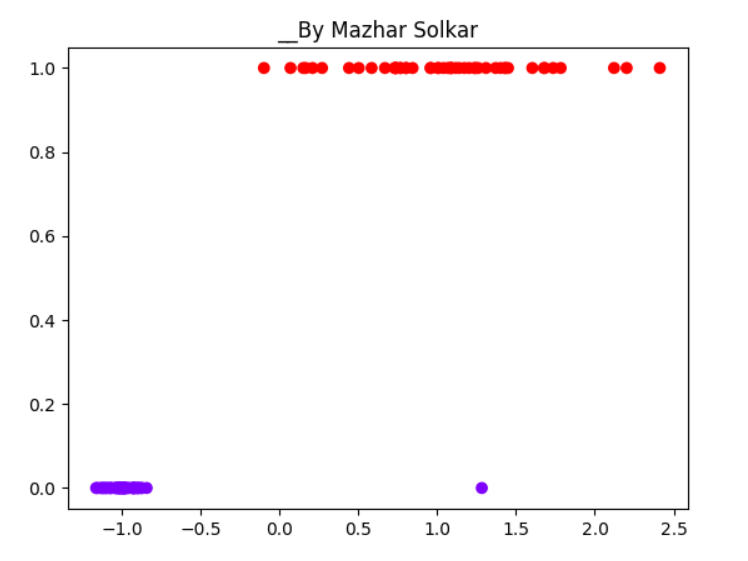
LogisticRegression() :-

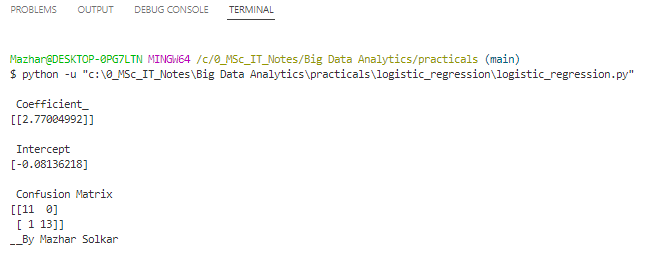
This method is used to implement logistic regression.

**Program :-**



**Output :-**





**Practical No 4**

**Aim: Implement an application that stores big data in MongoDB and manipulate it using python.**

**Description :-**

**MongoDB**, the most popular NoSQL database, is an open-source document-oriented database. The term ‘NoSQL’ means ‘non-relational’. It means that MongoDB isn’t based on the table-like relational database structure but provides an altogether different mechanism for storage and retrieval of data.

SQL databases store data in tabular format. This data is stored in a predefined data model which is not very much flexible for today’s real-world highly growing applications. Modern applications are more networked, social and interactive than ever. Applications are storing more and more data are accessing it at higher rates.

Relational Database Management System(RDBMS) is not the correct choice when it comes to handling big data by the virtue of their design since they are not horizontally scalable. If the database runs on a single server, then it will reach a scaling limit. NOSQL databases are more scalable and provide superior performance. MongoDB is such a NoSQL database that scales by adding more and more servers and increases productivity with its flexible document model.

**Methods :-**

1. MongoClient('localhost:27017') :-

This method is used to get at which port monodb is running.

2. client.get\_database('database\_name') :-

This method is used to access the database.

3. db.records\_name :-

This method is used to access the collection of database.

4. records.count\_documents({}) :-

This method is used to count the number of records in the collection.

5. list(records.find()) :-

This method is used to print all the records in collection.

6. records.update\_one({"$set":{"key","value"}}) :-

This method is used to update one record in collection.

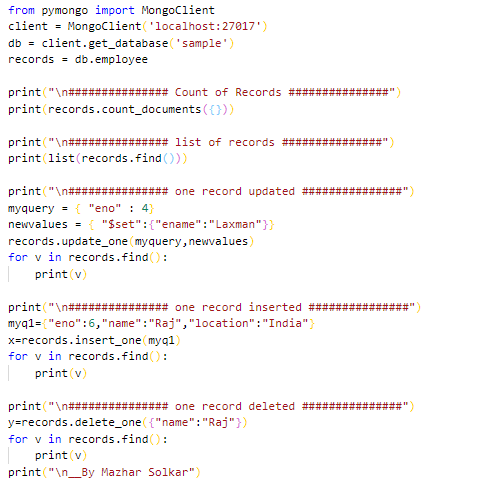
7. records.insert\_one({"eno":6,"name":"Raj","location":"India"}) :-

This method is used to insert one record in collection.

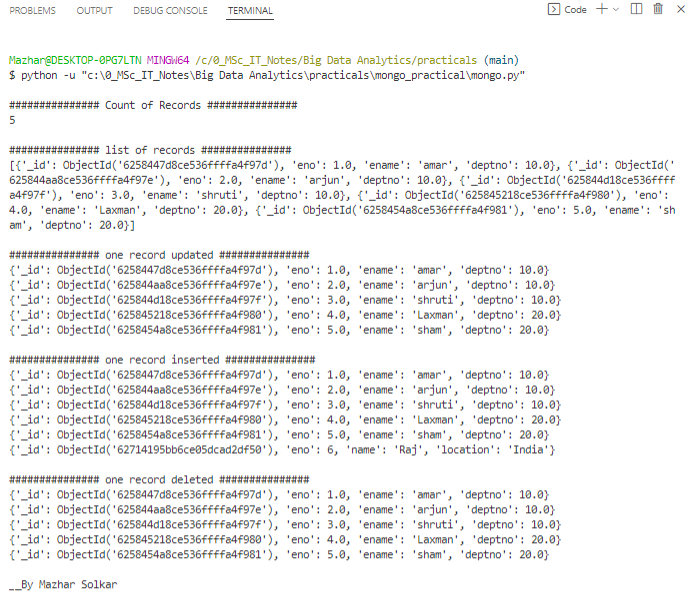
8. records.delete\_one({"name":"Raj"}) :-

This method is used to delete one record from collection.

**Program :-**



**Output :-**



**Practical No 5**

**Aim: Implement SVM classification Technique.**

**Description :-**

SVM is a famous supervised machine learning algorithm used for classification as well as regression algorithms. However,  mostly it is preferred for classification algorithms. It basically separates different target classes in a hyperplane in n-dimensional or multidimensional space.

The main motive of the SVM is to create the best decision boundary that can separate two or more classes(with maximum margin) so that we can correctly put new data points in the correct class. Because It chooses extreme vectors or support vectors to create the hyperplane, that’s why it is named so.

**Methods :-**

1. StandardScaler() :-

It is used for feature scaling.

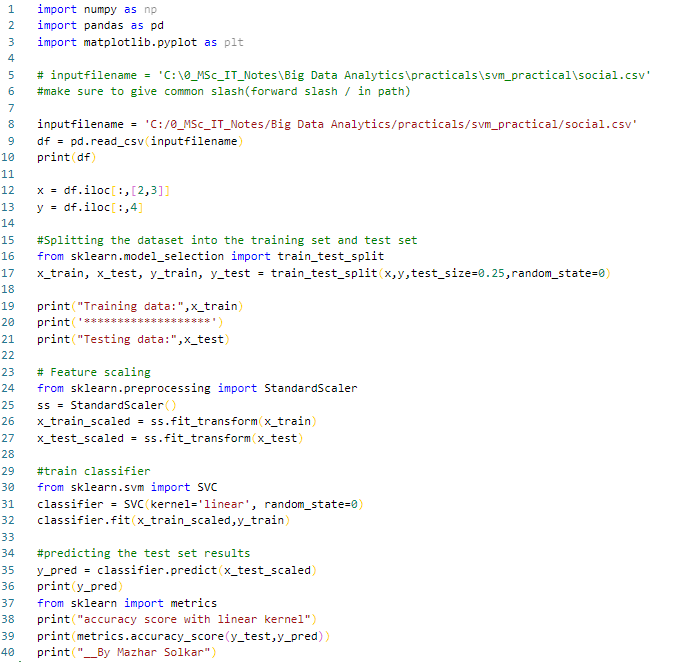
2. SVC(kernel='linear', random\_state=0) :-

This method is used for implementing SVM.

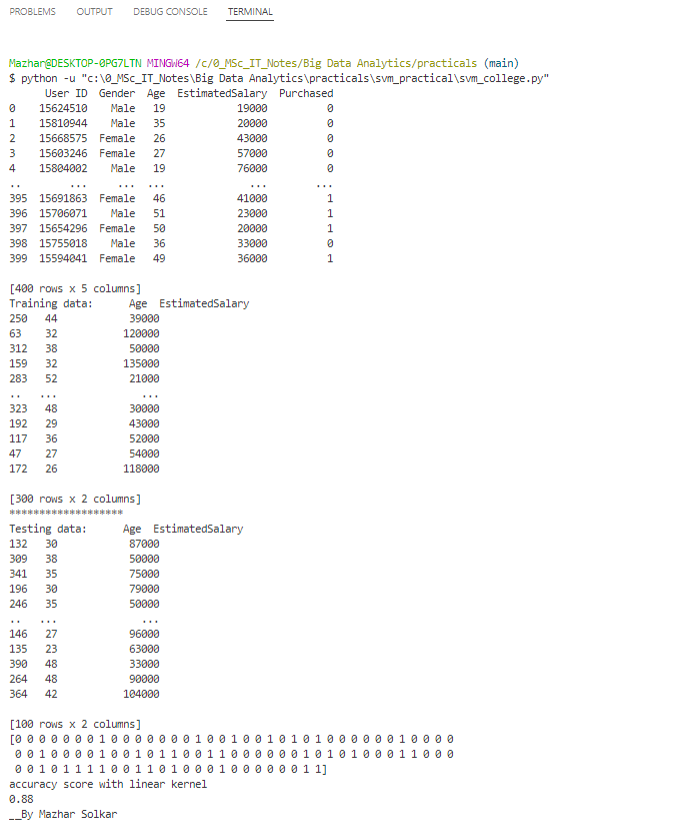
3. metrics.accuracy\_score(y\_test,y\_pred) :-

This method is used to check the accuracy score of the model.

**Program :-**



**Output :-**



**Practical No 6**

**Aim: Implement Decision Tree classification Technique.**

**Description :-**

Decision Tree is a supervised learning method used in data mining for classification and regression methods. It is a tree that helps us in decision-making purposes. The decision tree creates classification or regression models as a tree structure. It separates a data set into smaller subsets, and at the same time, the decision tree is steadily developed. The final tree is a tree with the decision nodes and leaf nodes. A decision node has at least two branches. The leaf nodes show a classification or decision. We can't accomplish more split on leaf nodes-The uppermost decision node in a tree that relates to the best predictor called the root node. Decision trees can deal with both categorical and numerical data.

**Methods :-**

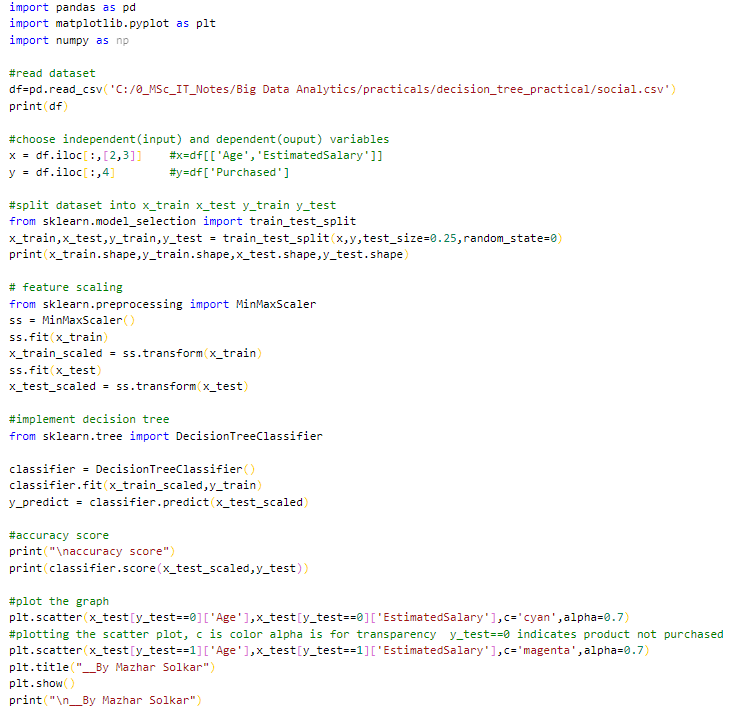
1. MinMaxScaler() :-

This method is used for feature scaling.

2. DecisionTreeClassifier() :-

This method is used to implement decision tree

**Program :-**



**Output :-**

